Distributed Optimal Traffic Lights Design for Large-Scale Urban Networks

Pietro Grandinetti, Carlos Canudas de Wit, Federica Garin

Distributed optimal traffic lights design for large-scale urban networks

Pietro Grandinetti, Carlos Canudas de Wit, *Fellow, IEEE*, and Federica Garin *Member, IEEE*

***Abstract*— In this paper we deal with the problem of dynamical assignment of traffic lights schedules in large-scale urban networks. We present a model for signalized traffic networks, based on the Cell Transmission Model, and then a simplified model based on averaging theory. The control objective is to improve traffic, optimizing traffic indexes such as total travel distance and density balancing. We design a scheme that decides the duty cycles of traffic lights, by solving a convex program. The optimization is done in real time, at each cycle of traffic lights, so as to take into account variable traffic demands. The scalability problem is tackled through the synthesis of a distributed optimization algorithm; this reduces the computational load significantly, since the large optimization problem is broken into small local subproblems, whose size does not grow with the size of the network, together with iterative exchanges of messages with few neighbor subproblems. The perfomance of the proposed approach is evaluated via numerical simulations in two different scenarios: a macroscopic (MatLab-based) Manhattan grid and a microscopic scenario (based on Aimsun simulator) reproducing a portion of the city of Grenoble, France.**

I. Introduction

Traffic congestion has always been a crucial problem for the design of efficient infrastructures, but it is particularly in the second half of the last century that this phenomenon has become predominant, due to the quickly increasing traffic demand and the more frequent congestions. Congestions appear when the amount of vehicles trying to use a common transportation route exceeds the latter's capacity. When they arise, they may lead to queueing phenomena or, even worse, to a severe degradation of the available infrastructure's usage. Hence, congestions result in reduced safety, increased pollution and excessive delays. Recent studies on the economic implication of such scenarios are [1], [2].

To be effective, strategies to address the congestion problem must have many important properties. They must be able to improve traffic behavior, e.g., optimizing the fluidity of traffic within cities. They must be adaptive, i.e., smartly reacting to changes in the local traffic demand. The optimization must be numerically tractable, to allow real-time implementation, leading to a trade-off between quality of the solution and computation time. A distributed implementation is also needed, both to achieve scalability, with a computational load not increasing with the network size, and to obtain resilience against failures, avoiding to have a single crucial centralized point.

P. Grandinetti is with Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, GIPSA-Lab, 38000 Grenoble, France (e-mail: pietro.grandinetti@gipsa-lab.fr)

C. Canudas-de-Wit is with CNRS, GIPSA-lab, 38000 Grenoble, France (e-mail: carlos.canudas-de-wit@gipsa-lab.fr)

F. Garin is with Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, GIPSA-Lab, 38000 Grenoble, France (e-mail: federica.garin@inria.fr)

Control techniques for urban infrastructures fall into three categories: control of signals offset (see, e.g., [3]) typically aims at producing green-wave effects, even though those are difficult to achieve in high load scenarios; routing policies, that have been developed in both centralized [4] and distributed [5] way, and that need technologies with high penetration rate in order to have a real impact; green time control for traffic lights, an intensively studied problem that is also the focus of our work.

The original traffic light policies were fixed-time techniques, the most famous example of which being [6]. Such techniques use predefined stages and minimum green times, and use hill climbing algorithms to find values of green times that minimize the number of stops. The drawback is that they heavily rely on historical rather than real time data. On the contrary, responsive techniques try to adapt their decision process to the current state. SCOOT [7] is a very popular adaptive strategy, built on the TRANSYT method [6]. SCOOT includes a central network model, fed with real measurements, that is in charge of evaluating changes in traffic; when some change turns out to be an improvement, it is sent to a local controller. Recently, also algorithms based on optimal control have been developed. In [8], [9] a quadratic program based on the store-and-forward model is used to produce values of green times that balance the queues at intersections. Also the max-pressure approach (see, e.g., [10]) uses a store-and-forward model where, however, queues at intersections have unlimited queue-length; under this assumption, max-pressure is proven to maximize the troughput by stabilizing the network. A more recent version of max-pressure is given in [11], where the cycle duration and the order of phases are imposed as constraints and only the duration of phases is optimized.

The contribution of this paper is to synthesize an optimal urban traffic lights control that satisfies the above-mentioned requirements. Our study includes modeling, formulating an optimization problem, finding a scalable distributed solution, and finally showing the potential impact with realistic simulations.

We start off with the Cell Transmission Model (CTM) [12], a widely used model in traffic applications and control, that is based on the demand-supply paradigm and therefore avoids any possibility of unlimited queue length. The CTM has been already employed in several works, both on modeling and control, for example in [3], [13] for arterials control via variable speed limits and ramp metering, in [4], [5], [14] to study stabilization of road networks, in [15], [16], [17] for freeway control via ramp metering, in [18], [19] as a model from which to derive a further simplified model, and in [8], [9], [20] for model predictive control.

In this work we formulate a variant of the CTM that

represents urban traffic and includes traffic lights signals by making use of the average theory ([21], [19]). We take into account standard urban traffic measures [15], [16], that are meant to improve traffic fluidity, and we consider the corresponding multi-objective optimization problem. By considering a one-step-ahead horizon for the optimization, we obtain a problem that we can turn into a convex program, and hence can be solved efficiently. Furthermore, we extend this result by developing a distributed control algorithm that in addition satisfies the requirement of scalability, exploiting the dual decomposition technique [22], [23]. We prove that, for the selected traffic cost function which includes total travel distance, density balancing and traffic lights regularization, the proposed distributed algorithm converges to the global optimum. Finally, we provide a large set of simulations, in which we test our method not only in a macroscopic simulated traffic network, but also in a commercial microscopic traffic simulator (Aimsun [24]), with realistic scenarios of Grenoble downtown traffic.

II. Traffic networks model

We start this section introducing the *signalized cell transmission model*, i.e., a model for traffic evolution that is based on the cell transmission model and incorporates traffic signals at intersections. Then, we formulate an averaged version, that we call *averaged cell transmission model*, which gives a simplified description, useful for control purposes. We show results on the numerical comparison between the two models, and close the section with the discrete-time version used in the remainder of the paper.

A. The signalized cell transmission model

To describe traffic evolution we use the description given by a mass conservation law [25], [26] and by its representation widely known as cell transmission model (CTM) [12]. In particular, we consider the extension of the CTM for networks with First In First Out (FIFO) policy at intersections that was proposed by Daganzo in [27].

An urban network is a collection $\mathcal{R}$ of roads, among which there are entering roads ($\mathcal{R}^{\text{in}}$) and exiting roads ($\mathcal{R}^{\text{out}}$). For each road i, the out-going traffic is regulated by a traffic light $u_i(t)$, alternating a *green* phase ($u_i(t) = 1$) and a *red* phase ($u_i(t) = 0$) within a cycle time of length T. We refer to *intersections* as the locations where two or more roads merge. Such intersections have no capacity storage. When two roads i and j are connected to the same intersection in such a way that flow can exit i and enter j, we say that i belongs to the set of upstream neighbors of j, a set indicated with $\mathcal{N}_j^-$. The set of downstream neighbors of j, $\mathcal{N}_j^+$, is defined accordingly. Furthermore, to every pair of roads i, j is associated a value $\beta_{ij} \in [0, 1]$ (split ratio)[1] that expresses the percentage of the flow exiting i which wants to turn in j ($\beta_{ij} = 0$ if and only if $j \notin \mathcal{N}_i^-$ and $\beta_{ij} = 1$ if and only if $\mathcal{N}_i^+ = \{j\}$).

[1]More precisely, β_{ij}'s are average split ratios. In our model, we disregard unknown random variations of the split ratios around their average. The further assumption that average split ratios are constant is made since their variation is slower than traffic dynamics, but time-varying $\beta_{ij}(t)$ can be introduced in our formalism, throughout the paper.

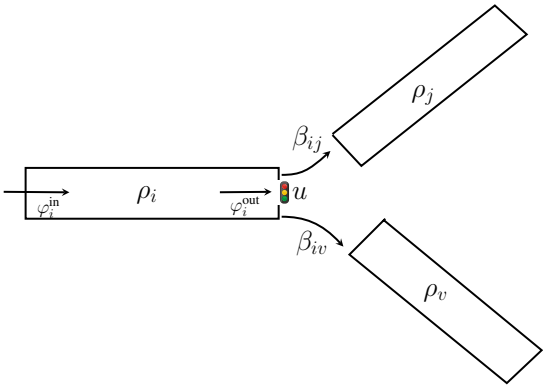


Fig. 1: Illustration of the notation used in the traffic model.

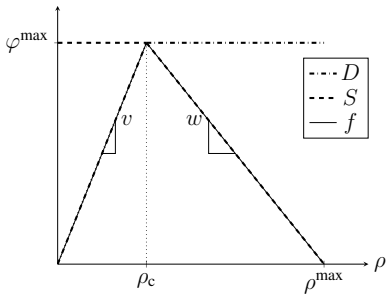


Fig. 2: Graphical representation of demand and supply functions (D and S) and of the triangular fundamental diagram $f = \min(D, S)$.

In urban scenarios we consider every road i as a cell of the CTM. The density of vehicles of a road i, $\rho_i(t)$, depends on the flows φ_i^{in} and φ_i^{out}, entering and exiting i, respectively (see Figure 1 for an illustration). Inflow and outflow values are given according to the demand and supply paradigm [28]. For a road i, its demand D_i is the flow of vehicles that want to exit i; its supply S_i is the flow that i can receive according to its storage capacity. We assume that D_i and S_i are the piecewise affine functions depicted in Figure 2.

Every road i is characterized by given parameters: the maximum speed in free flow v_i, the speed in congestion phase w_i, the maximum density $\rho_i^{\max}$, the maximum flow $\varphi_i^{\max}$ and the length of the road L_i.

Finally, external traffic demand for every road i entering the network is indicated by D_i^{in}, and external supply for every road j exiting the network is indicated by S_j^{out}. Thus, summarizing, the signalized traffic model is given by the following set of equations:

$$\dot{\rho}_i(t) = \frac{1}{L_i}\left(\varphi_i^{\text{in}}(t) - u_i(t)\varphi_i^{\text{out}}(t)\right), \tag{1a}$$

$$\varphi_i^{\text{out}}(t) = \begin{cases} \min\{D_i(t), S_i^{\text{out}}(t)\}, i \in \mathcal{R}^{\text{out}} \\ \min\left\{D_i(t), \left\{\frac{S_j(t)}{\beta_{ij}}\right\}_{j\in\mathcal{N}_i^+}\right\}, \text{otherwise} \end{cases} \tag{1b}$$

$$\varphi_i^{\text{in}}(t) = \begin{cases} \min\{D_i^{\text{in}}(t), S_i(t)\}, i \in \mathcal{R}^{\text{in}} \\ \sum_{j\in\mathcal{N}_i^-} u_j(t)\beta_{ji}\varphi_j^{\text{out}}(t), \text{otherwise} \end{cases} \tag{1c}$$

$$D_i(t) = \min\{v_i\rho_i(t), \varphi_i^{\max}\}, \tag{1d}$$

$$S_i(t) = \min\{\varphi_i^{\max}, w_i(\rho_i^{\max} - \rho_i(t))\}. \tag{1e}$$

In model (1) every traffic light is naturally constrained to be either 0 or 1. Moreover, in every intersection, a *collision avoidance* constraint must hold, i.e., at every time instant at most one traffic light can be green among the ones connected to the same intersection; this contraint is expressed as follows:

$$\forall\, i \in \mathcal{R} \setminus \mathcal{R}^{\text{in}},\, \forall\, t \geq 0,\ \sum_{q \in \mathcal{N}_i^-} u_q(t) \leq 1. \tag{2}$$

B. The averaged cell transmission model

The main difficulty with the model given in the previous section is due to its complexity: traffic lights are represented by binary variables (i.e., differentiating the *green* from the *red*). Therefore, working with this model is usually hard, for example, algorithms meant to schedule optimal values for traffic lights are in most cases *mixed-integer programs* (see, e.g., [29], [30]), known to be intractable.

Our approach in [31] was to define an averaged model: since the typical behavior of traffic lights is to alternate green and red within a given cycle time, our model considers, as controlled variables, the *duty cycle* of every traffic light, i.e., the percentage of green time in a cycle, formally defined as $\bar{u}_i(t) = \frac{1}{T}\int_{kT}^{(k+1)T} u_i(t)\,\mathrm{d}t$, $k \in \mathbb{N}$. Notice that, while $u(t)$ was a binary valued signal, $\bar{u}(t)$ is a signal that takes a constant (average) real value through an entire cycle.

The model we propose, with respect to other models based on similar ideas, e.g., the store and forward [8], is more faithful to the original CTM, in the sense that it relies on the concept of demand and supply, and is able to reproduce congestion spillbacks[2] . Densities of the averaged model imitate densities of the signalized CTM: their time evolution is described by the same equation, except that the binary variables describing traffic lights are replaced by the corresponding duty cycles, as follows.

$$\dot{\bar{\rho}}_i(t) = \frac{1}{L_i}\left(\varphi_i^{\text{in}}(t) - \bar{u}_i(t)\varphi_i^{\text{out}}(t)\right), \tag{3a}$$

$$\varphi_i^{\text{out}}(t) = \begin{cases} \min\{D_i(t), S_i^{\text{out}}(t)\}, i \in \mathcal{R}^{\text{out}} \\ \min\left\{D_i(t), \left\{\frac{S_j(t)}{\beta_{ij}}\right\}_{j\in\mathcal{N}_i^+}\right\}, \text{otherwise} \end{cases} \tag{3b}$$

$$\varphi_i^{\text{in}}(t) = \begin{cases} \min\{D_i^{\text{in}}(t), S_i(t)\}, i \in \mathcal{R}^{\text{in}} \\ \sum_{j\in\mathcal{N}_i^-} \bar{u}_j(t)\beta_{ji}\varphi_j^{\text{out}}(t), \text{otherwise} \end{cases} \tag{3c}$$

$$D_i(t) = \min\{v_i\bar{\rho}_i(t), \varphi_i^{\max}(t)\}, \tag{3d}$$

$$S_i(t) = \min\{\varphi_i^{\max}(t), w_i(\rho_i^{\max} - \bar{\rho}_i(t))\}. \tag{3e}$$

In the averaged system (3), duty cycles are constrained to have real values between 0 and 1. Moreover, the collision avoidance condition (2) now constrains the sum of the duty cycles, i.e.,

$$\forall\, i \in \mathcal{R} \setminus \mathcal{R}^{\text{in}},\, \forall\, t \geq 0\ \sum_{q \in \mathcal{N}_i^-} \bar{u}_q(t) \leq 1. \tag{4}$$

[2]A different approach is proposed in [32], where the store-and-forward model is enriched in such a way to reproduce congestion spillbacks.

The state variable $\bar{\rho}$, solution of (3), as it turns out, is a good approximation of the state variable ρ solving (1). This property, while often used in model-based control approaches, has received little theoretical attention. Relevant exceptions are [18], [19]. We recall here the main result about the comparison between the dynamical systems (3) and (1); it provides a bound on the error $|\bar{\rho} - \rho|$, which holds for a single road and under suitable technical assumptions discussed in [19].

Proposition 1 ([19, Theorem 2])**.** *Given $\varepsilon = 1/L$, there exists a constant $C(T)$ (proportional to the cycle length T), such that*

$$|\rho(t) - \bar{\rho}(t)| \leq C(T)\,\varepsilon,$$

for every $t \geq 0$.

In Sect. II-C, we present simulations, illustrating the error $|\bar{\rho} - \rho|$.

C. Numerical comparison

For our software simulations, we consider a Manhattan-like grid, namely a network of horizontal and vertical roads connected by standard 4-ways intersections as depicted in Figure 3a. We assume that all roads have the same parameters: for all i, $\varphi_i^{\max} = \varphi^{\max} = 2000$ vehicles per hour (veh/h), $\rho_i^{\max} = \rho^{\max} = 200$ vehicles per km (veh/km), $v_i = v = 50$ km/h, $w_i = w = 12.5$ km/h, $L_i = L = 0.5$ km. Notice that these values imply that the critical density is $\rho_c = 40$ veh/km. Split ratios are chosen as follows. If a road i has two downstream neighbors, then the neighbor along the same (horizontal or vertical) orientation as i is assigned a split ratio of 0.6 plus a small random number, extracted uniformly in $[-0.05, 0.05]$; the split ratio for the second downstream road is then assigned so that the two sum to one. If there is only one downstream neighbor, then it gets split ratio 1. The simulations are run for 720 time steps. Until the 550th step, an external demand requires to enter the network: on all entering roads, at each time instant, there is an incoming flow uniformly random in the interval $[\varphi^{\max}/2, \varphi^{\max}]$. Then, the external demand is set to zero. By doing so, we reproduce both the case where the network is filling up and the one where it is emptying.

In this section, for ease of visualization of the results, we consider a network size of 40 roads, i.e., our network is the graph shown in Figure 3a.

We are interested in evaluating:

- how well the averaged system reproduces the distinction between free-flow roads (i.e., with density lower than the critical density) and congested roads that is given by the signalized system;
- how close the averaged system (3) is to the signalized CTM (1).

Regarding the second point, we compare the time trajectories of the two systems, and also of the formal integral average of the signalized CTM trajectory, defined as [21]:

$$\hat{\rho}(t) = \frac{1}{T}\int_t^{t+T} \rho(\tau)\,\mathrm{d}\tau. \tag{5}$$

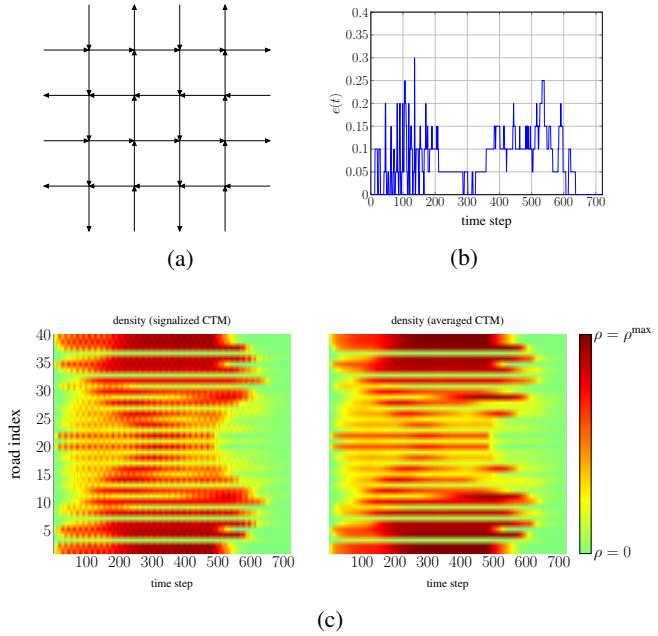


Fig. 3: Results of numerical validation regarding the accuracy of the averaged CTM. Figure 3a shows the idealized network of a Manhattan-like grid. We compare the evolution of an average CTM and a signalized CTM subject to the same external demands and supply. Every road can have a status either free or congested: in Figure 3b is depicted the percentage of status that are not correct (in the sense that in the average CTM they differ from the signalized CTM), as $e(t) = \frac{\#\text{wrong status}}{\#\text{roads}}$. In Figure 3c is depicted the time evolution of densities for all roads in the network; the figure shows that the averaged densities are very close to the signalized densities.

We remark that, due to the nonlinearity in the expression of $\rho(t)$ in (1), the value of $\bar{\rho}(t)$ is not exactly the same as $\hat{\rho}(t)$, but it is a good approximation of the latter.

Representative examples of our results are reported in Figure 3. Notice that the averaged system succeeds in discriminating the status (free or congested) of a high percentage of roads. The mean error is around 10%, as it results from Figure 3b. Figure 3c compares the densities obtained with signalized CTM and with the averaged model. We can see that they are very similar. We also notice that the averaged model smooths out the small rapid changes that appear with the signalized model, as expected, since $\bar{\rho}(t)$ is rather meant to approximate the integral average $\hat{\rho}(t)$, and not the instantaneous value $\rho(t)$. This is illustrated in Figure 4, where we show the density trajectory for a single selected road, comparing the densities $\rho(t)$, $\hat{\rho}(t)$ and $\bar{\rho}(t)$ obtained with (1), (5), and (3), respectively.

More results about quantitative comparisons are given in Table I. We run simulations with varying length of the traffic lights cycle and we measure the errors $|\bar{\rho}_i(t) - \hat{\rho}_i(t)|$ and $|\bar{\rho}_i(t) - \rho_i(t)|$, where $\bar{\rho}_i$, $\hat{\rho}_i$ and ρ_i are obtained from (3), (5), and (1). Table I reports mean and worse-case errors, across all time steps and all roads. Such values should be interpreted recalling that densities vary in the interval $[0, \rho^{\max}]$, and here $\rho^{\max} = 200$ veh/h. We can see that the mean error is very low, and the worse-case error remains acceptable.

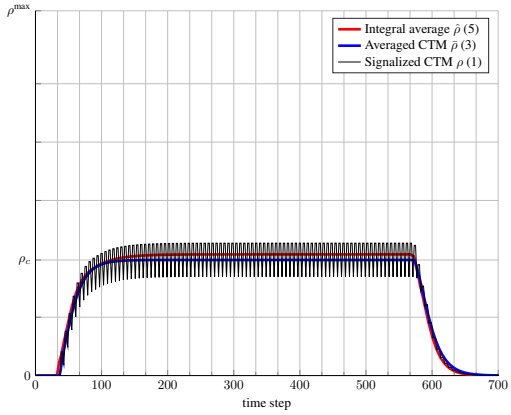


Fig. 4: Example of comparison: for one selected road i, evolution of densities ρ_i, $\bar{\rho}_i$, and $\hat{\rho}_i$, obtained from (1), (3), and (5), respectively.

TABLE I: Validation with different cycle lengths.

		$T = 45$ s	$T = 60$ s	$T = 90$ s	$T = 120$ s
Error (3) vs (5) veh/km	Mean	1.9	2.03	1.69	1.8
	Worst	12	12	10	11
Error (3) vs (1) veh/km	Mean	2.5	2.7	3	4.7
	Worst	13	15	14	22

D. Discrete-time model

The purpose of having the model (3) is to ease the synthesis of a control policy for the traffic lights by transforming binary control signals into continuous ones. In addition, for practical use of the model, a discrete-time version of (3) is needed. We discretize the system using the Euler discretization method, with sampling time T_s. The choice of the sampling time must satisfy the requirement stressed in Daganzo's seminal work [12], i.e., $v_i T_s / L_i < 1$, for every road i, in order to have a stable model (the eigenvalues of the discrete-time system have the form of the left-hand side of the previous inequality). A more in-depth discussion about such a choice is provided in [33].

In conclusion, the model we will use for the optimal control strategy is given by the following equations:

$$\bar{\rho}_i(t + T_s) = \bar{\rho}_i(t) + \frac{T_s}{L_i}\left(\varphi_i^{\text{in}}(t) - \bar{u}_i(t)\varphi_i^{\text{out}}(t)\right), \tag{6a}$$

$$\varphi_i^{\text{out}}(t) = \begin{cases} \min\{D_i(t), S_i^{\text{out}}(t)\}, i \in \mathcal{R}^{\text{out}} \\ \min\left\{D_i(t), \left\{\frac{S_j(t)}{\beta_{ij}}\right\}_{j \in \mathcal{N}_i^+}\right\}, \text{otherwise} \end{cases} \tag{6b}$$

$$\varphi_i^{\text{in}}(t) = \begin{cases} \min\{D_i^{\text{in}}(t), S_i(t)\}, i \in \mathcal{R}^{\text{in}} \\ \sum_{j \in \mathcal{N}_i^-} \bar{u}_j(t)\beta_{ji}\varphi_j^{\text{out}}(t), \text{otherwise} \end{cases} \tag{6c}$$

$$D_i(t) = \min\{v_i \bar{\rho}_i(t), \varphi_i^{\max}\}, \tag{6d}$$

$$S_i(t) = \min\{\varphi_i^{\max}, w_i(\rho_i^{\max} - \bar{\rho}_i(t))\}. \tag{6e}$$

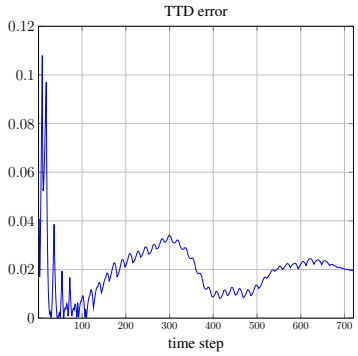


Fig. 5: Validation of the averaged model regarding total travel distance. The data plotted are computed as $\frac{|\text{TTD}(t)_{\text{signalized}} - \text{TTD}(t)_{\text{averaged}}|}{\text{TTD}(t)_{\text{signalized}}}$

III. One-step-ahead centralized optimal traffic control

In this section, we start by recalling the definition of a well known urban traffic performance measure, the total travel distance. We continue by showing that the average CTM also succeds in approximating such an index reasonably well, that is a further indication of the usefulness of this model for control purposes. We then formulate the traffic control problem, and proceed to show how this problem can be treated in the framework of efficient optimization.

A. Urban traffic measures and problem formulation

1) Total travel distance: A standard metric in traffic systems is the Total Travel Distance (TTD) [15], a cumulative index defined as

$$\text{TTD}(t) = \sum_{k=0}^{\lfloor t/T_s \rfloor} \sum_{i \in \mathcal{R}} f_i(kT_s), \tag{7}$$

where $f_i(\tau)$ is the flow inside the road i at time τ, given by

$$f_i(\tau) = \min\left\{v_i \rho_i(\tau), w_i(\rho_i^{\max} - \rho_i(\tau))\right\}.$$

TTD is a measure that should be maximized, as it increases the flow travelling the section.

It is worth mentioning that the averaged CTM offers good approximation also for the TTD. While performing simulations illustrated in Section II-C we also measure the TTD for the averaged and the signalized systems. The result is reported in Figure 5, and we notice that error is lower than 10%, and lower than 4% almost always. This means that the averaged CTM can consistently be used to predict values of TTD and density, and therefore used in a model based optimization scheme.

2) Density balancing: Balancing was introduced in traffic scenarios in [17]. The main idea is to try to evenly use the infrastructure, avoiding to have some overly congested roads and some underutilized others. Moreover, balanced density can be regarded as an equal inter-distance between vehicles, which makes travels smoother and safer, and reduces emissions [16].

A measure of densities balancing is given by

$$\text{Bal}(t) = \sum_{k=0}^{\lfloor t/T_s \rfloor} \sum_{i \in \mathcal{R}} \sum_{j \in \mathcal{N}_i^+} (\bar{\rho}_i(kT_s) - \bar{\rho}_j(kT_s))^2. \tag{8}$$

Equation (8) can be expressed in a compact form as

$$\text{Bal}(t) = \sum_{k=0}^{\lfloor t/T_s \rfloor} \bar{\rho}'(kT_s) \mathcal{L} \bar{\rho}(kT_s), \tag{9}$$

where $'$ denotes transpose, and $\mathcal{L}$ is a laplacian matrix of the network, defined as

$$\begin{aligned} \mathcal{L}_{ii} &= |\mathcal{N}_i^- \cup \mathcal{N}_i^+|, \\ \mathcal{L}_{ij} &= \begin{cases} -1 & \text{if } j \in \mathcal{N}_i^- \cup \mathcal{N}_i^+ \\ 0 & \text{elsewhere.} \end{cases} \end{aligned} \tag{10}$$

3) Traffic lights regularization: The structure of the CTM with FIFO rule is such that numerical optimization can lead to abrupt changes in the control dynamics, even for short-term disturbances on the system input. This issue is commonly addressed in optimization by adding a regularization term to the cost function. Regularizing means penalizing abrupt variations of the decision variables in order to obtain smoother results across time. For the one-step-ahead traffic control this translates into adding the following term to the cost function,

$$\|\bar{u}(t) - \bar{u}(t-T)\|_2^2, \tag{11}$$

where $\bar{u}(t)$ is the duty cycle for the upcoming cycle, as defined earlier, and $\bar{u}(t-T)$ is the duty cycle applied in the last cycle.

4) Problem formulation: A control policy intended to improve urban traffic behavior according to above-discussed criteria may be verbalized as: find duty cycles values to be applied in the signalized intersections, such that they optimize measures (7), (8) and (11), given the current state of the network (density). Duty cycles are constrained to take values lower than 1 and larger than assigned lower bounds $l_i \geq 0$, and to respect the collision avoidance constraint (4). Notice that the inequality in the constraint permits some time instants where all lights are red; for instance, if there is a strong congestion downstream it may be useful to stop vehicles upstream. We use positive constant values k_{bal} and k_{ttd} to weigh the TTD and the density balancing in the multi-objective optimization.

Our control scheme is the following: at time t, when a traffic lights cycle begins, novel values for the duty cycles are computed by solving the following optimization problem, over a time horizon of K steps:

$$\begin{aligned} \min_{\bar{u}} \quad & \sum_{k=1}^{K} \Bigg(k_{\text{bal}} \bar{\rho}'(t+kT_s) \mathcal{L} \bar{\rho}(t+kT_s) \\ & \qquad - k_{\text{ttd}} \sum_{i \in \mathcal{R}} f_i(t+kT_s) \Bigg) \\ & \qquad + \|\bar{u} - \bar{u}(t-T)\|_2^2 \\ \text{subject to (6) and to} \quad & \\ l_i \leq \bar{u}_i \leq 1, \quad & \sum_{j \in \mathcal{N}_i^-} \bar{u}_j \leq 1, \forall\, i, \end{aligned} \tag{P1}$$

where densities $\bar{\rho}_i(t)$ are the densities measured at time t. Notice that the solution to program (P1) is one value for every traffic light. Such values are then applied into the network for

the entire upcoming cycle; then, this process is repeated at the beginning of every cycle.

In the problem (P1) we aim at optimizing a cost function given by the model-based predictions over some time horizon and the entire network. However, in order to achieve an opimization problem that is computationally tractable, a trade-off is required. We will therefore limit the predictions computed at time t to one step ahead only, choosing $K = 1$, i.e., to values depending on $\bar{\rho}(t + T_s)$.

B. A convex formulation of the control problem

We will now show that a convex program can be synthesized, if the optimization is limited to one step ahead in prediction.

First, it is worth noticing that every term $\varphi_i^{\text{out}}(t)$ is a real number, computed using the measured values $\bar{\rho}_j(t)$'s. Hence, from (6), it is clear that the relation between the one-step-ahead density $\bar{\rho}(t + T_s)$ and the duty cycles $\bar{u}$ is affine. We can write it as

$$\bar{\rho}_i(t + T_s) = h_i'(t)\bar{u}_{[i,v\in\mathcal{N}_i^-]} + \bar{\rho}_i(t), \tag{12}$$

where $\bar{u}_{[i,v\in\mathcal{N}_i^-]}$ indicates a vector containing duty cycles of road i and roads $v \in \mathcal{N}_i^-$, and $h_i(t)$ is a vector of real numbers and consistent dimension. In matrix form, the whole vector of one-step-ahead densities depends affinely from the whole vector of duty cycles, as follows

$$\bar{\rho}'(t + T_s) = H(t)\bar{u} + c(t),$$

where the rows of matrix H are the vectors h_i''s. With this, we can consider the following portion of the cost function, with the balancing and regularization terms,

$$k_{\text{bal}}\bar{\rho}'(t + T_s)\mathcal{L}\bar{\rho}(t + T_s) + \|\bar{u} - \bar{u}(t - T)\|_2^2,$$

and we can re-write it as the following quadratic function of $\bar{u}$ only

$$\bar{u}'Q\bar{u} + p'\bar{u},$$

where

$$Q = k_{\text{bal}}H'(t)\mathcal{L}H(t) + I, \tag{13a}$$

$$p = 2k_{\text{bal}}H'(t)\mathcal{L}c(t) - 2\bar{u}(t - T). \tag{13b}$$

Next we look at the TTD. By substituting (3b) in (3c) and then into (3a), every element of the cost function becomes a nonlinear function of the duty cycles, expressed as

$$f_i(\bar{u}) = \min\left\{ v_i\left(\bar{\rho}_i(t) + \frac{T_s}{L_i}\Big(\sum_{j\in\mathcal{N}_i^-}\bar{u}_j\beta_{ji}\varphi_j^{\text{out}} - \bar{u}_i\varphi_i^{\text{out}}\Big)\right), \right.$$
$$\left. w_i\left(\rho_i^{\max} - \bar{\rho}_i(t) - \frac{T_s}{L_i}\Big(\sum_{j\in\mathcal{N}_i^-}\bar{u}_j\beta_{ji}\varphi_j^{\text{out}} + \bar{u}_i\varphi_i^{\text{out}}\Big)\right)\right\}. \tag{14}$$

Rewriting $\text{TTD} = \mathbf{1}'f(\bar{u})$, where f is the vector containing all f_i's and $\mathbf{1}$ is the all-ones vector, we obtain the following optimization problem:

$$\begin{aligned}\min_{\bar{u}\geq 0}\quad & \bar{u}'Q\bar{u} + p'\bar{u} - k_{\text{ttd}}\mathbf{1}'f(\bar{u})\\ & \text{subject to, } \forall i\\ & l_i \leq \bar{u}_i \leq 1,\\ & \sum_{j\in\mathcal{N}_i^-}\bar{u}_j \leq 1.\end{aligned} \tag{P2}$$

The nonlinearity in the objective function of (P2) can be moved onto the constraints by defining a set of auxiliary variables y, along with the constraints $y_i = f_i(\bar{u}),\ \forall\, i$. Then such constraints, given their special structure due to the piecewise linearity of every f_i, can be relaxed, hence giving the following convex program.

$$\begin{aligned}\min_{\bar{u},y\geq 0}\quad & \bar{u}'Q\bar{u} + p'\bar{u} - k_{\text{ttd}}\mathbf{1}'y\\ & \text{subject to, } \forall i\\ & l_i \leq \bar{u}_i \leq 1\\ & \sum_{j\in\mathcal{N}_i^-}\bar{u}_j \leq 1\\ & y_i \leq v_i\left(\bar{\rho}_i(t) + \frac{T_s}{L_i}\Big(\sum_{j\in\mathcal{N}_i^-}\bar{u}_j\varphi_j^{\text{out}} - \bar{u}_i\varphi_i^{\text{out}}\Big)\right),\\ & y_i \leq w_i\left(\rho_i^{\max} - \bar{\rho}_i(t) - \frac{T_s}{L_i}\Big(\sum_{j\in\mathcal{N}_i^-}\bar{u}_j\varphi_j^{\text{out}} + \bar{u}_i\varphi_i^{\text{out}}\Big)\right).\end{aligned} \tag{P3}$$

We notice that (P3) is an instance of quadratic program with linear constraints, which belongs to the class of tractable problems and can be efficiently solved with Newton-like (interior point) methods, see [34]. Moreover, and most importantly, (P3) can be used to find an optimal solution of problem (P2), because the two turn out to be equivalent, as the following proposition shows.

Proposition 2. *Let Q and $D \in \mathbb{R}^{n\times n}$, and p, b, $g \in \mathbb{R}^n$, with $b \geq 0$. Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a concave positive function defined as $f(u) = \min\big(H_1u + c_1, H_2u + c_2\big)$, for given $H_1, H_2 \in \mathbb{R}^{n\times n}$ and $c_1, c_2 \in \mathbb{R}^n$.*

Let (15) *be the following optimization problem*

$$\min_{u\geq 0}\quad u'Qu + p'u - b'f(u) \tag{15a}$$

$$Du \leq g, \tag{15b}$$

and (16) *be the following one*

$$\min_{u,y\geq 0}\quad u'Qu + p'u - b'y \tag{16a}$$

$$Du \leq g \tag{16b}$$

$$y \leq H_1u + c_1 \tag{16c}$$

$$y \leq H_1u + c_2. \tag{16d}$$

Then (15) *is equivalent to* (16)*.*

Proof. The statement is true if, for all (u^*, y^*) optimal solutions of (16), $y^* = f(u^*)$ holds. By contradiction, suppose

(u^*, y^*) is an optimal solution for (16) and there exist $\varepsilon > 0$ and a natural $i \leq n$ such that

$$y_i^* + \varepsilon = \min(h_{1i}u^* + c_{1i}, h_{2i}u^* + c_{2i}),$$

where h_{1i} (h_{2i}) is the ith row of H_1 (H_2). Then, increasing the ith component of y^* by ε one obtains a new admissible solution that provides a better value of the objective function than (u^*, y^*), which, therefore, could not be optimal. □

Proposition 2 is a well-known technique to deal with some non convex constraint. In the context of traffic networks it has been employed for variable speed limit control and vehicle (split ratios) control, see [13], [14]. Here we use it in relation with the triangular shape of the fundamental diagram, for the synthesis of traffic lights duty cycles.

C. Remark on the cost function

In this paper we present a multiobjective optimization mainly directed to two well known traffic measures, the total travel distance and the density balancing, together with a regularization term. Notice that if one decides to include in the objective the TTD only, then it is possible to prove that the one-step-ahead optimization problem is a linear program, which on average is computationally even easier to solve than a quadratic. We refer the reader to our previous work [31] for more details.

The distributed algorithm presented in Section IV is suitable for the multiobjective optimization (P3). Indeed the convergence of the distributed scheme relies on the strict convexity of the objective function. However it is possible to obtain a distributed optimal algorithm for the linear program that considers TTD only. This can be done by using different algorithms, such as those proposed in [35], [36].

D. Remark on the computational details

In this section we clarify how the objective function of (P3) can be constructed from the measured densities. According to (12), $\bar{\rho}_i(t + T_s)$ depends only on decision variables $\bar{u}_v$ for $v \in \mathcal{N}_i^- \cup \{i\}$, and the dependency is affine. The computation of the coefficients in the linear combination (i.e., the entries of h_i) requires only the knowledge of φ_v^{out} for $v \in \mathcal{N}_i^- \cup \{i\}$. In turn, each φ_v^{out} is computed according to (6b), with the knowledge of densities $\bar{\rho}_w(t)$ only for $w \in \mathcal{N}_v^+ \cup \{v\}$. Figure 6 illustrates in an example the roads involved in the computation of h_i.

IV. Distributed control design

In this section we will develop a distributed version of the one-step-ahead optimal traffic control. A distributed control is a control technique in which a single central point of computation is missing, and the operations are distributed among several subproblems that can share information; in this way, requirements about efficiency and scalability are more likely to be satisfied. For the problem we are analyzing, the capability to synthesize a distributed algorithm is very relevant, since traffic networks often have very large dimension.

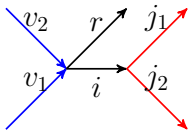


Fig. 6: Illustration of the roads involved in $h_i(t)$ in (12). Here $\mathcal{N}_i^- = \{v_1, v_2\}$ and $\mathcal{N}_i^+ = \{j_1, j_2\}$. According to (12), the one-step-ahead density of road i is an affine function of duty cycles $\bar{u}_i$, $\bar{u}_{v_1}$ and $\bar{u}_{v_2}$. The real numbers that are the coefficients of such a relation (the entries of $h_i(t)$) can be computed from $\varphi_i^{\text{out}}, \varphi_{v_1}^{\text{out}}, \varphi_{v_2}^{\text{out}}$. In turn, φ_i^{out} is obtained from $\bar{\rho}_{j_1}(t)$, $\bar{\rho}_{j_2}(t)$, and $\bar{\rho}_i(t)$, while $\varphi_{v_1}^{\text{out}}$, $\varphi_{v_2}^{\text{out}}$ are obtained from $\bar{\rho}_i(t)$ and $\bar{\rho}_r(t)$, together with $\bar{\rho}_{v_1}(t)$ and $\bar{\rho}_{v_2}(t)$, respectively.

We address this by designing an algorithm based on decomposing the original centralized optimization in subproblems of smaller dimension. In order to reconstruct the global optimal solution, the subproblems follow an iterative convergent scheme which includes some exchange of information. These exchanges take place according to a scheme, established a-priori, that is called communication graph. The advantage is that we can obtain subproblems having a fixed size, not growing with the network size, and solving them in parallel gives a scalable scheme.

We will first introduce the communication graph. Then, we will show how the centralized problem can be reformulated to allow a distributed algorithm to exist and work properly. Finally, we will present the distributed algorithm.

A. Communication graph

Our intent is to construct one subproblem for every road; in this way, each road is in charge of selecting a value for its own traffic light. However, subproblems are coupled, and each road needs some information from neighbor roads. The communication graph describes as neighbors the roads who exchange messages about variables.

We define the set $\mathcal{S}_i$ of roads that are allowed to share information with road i as

$$\mathcal{S}_i = \mathcal{N}_i^- \cup \mathcal{N}_i^+ \cup \mathcal{I}_i,$$

where $\mathcal{I}_i = \{q : \mathcal{N}_q^+ \equiv \mathcal{N}_i^+\}$.

It is worth noticing that the set $\mathcal{S}_i$ speficies connections between roads that are geographically close to each other, and therefore has some advantage also in view of a practical implementation. See Figure 7 for a graphical illustration.

The definition of $\mathcal{S}_i$ is instrumental to the distributed optimization formulated in the remainder of this section. Indeed, exchanged messages within neighborhood $\mathcal{S}_i$ can contain all the information needed for the formulation of the ith local subproblem, as discussed in the end of Sect. IV-B, as well as all the variables to be exchanged at each iteration of the algorithm presented in Sect. IV-C.

Finally, notice that $i \in \mathcal{S}_i$, and that $p \in \mathcal{S}_i$ implies $i \in \mathcal{S}_p$. Hence we can define the communication graph as the undirected graph $\mathcal{G}$ whose vertices are roads and in which an edge $\{i, j\}$ exists whenever $j \in \mathcal{S}_i$. The natural assumption that the directed road network is weakly connected implies that $\mathcal{G}$ is connected.

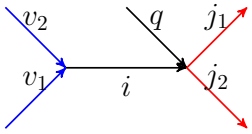


Fig. 7: Illustration of the communication neighborhood $\mathcal{S}_i$ defined in Sect. IV-A. It involves upstream roads in $\mathcal{N}_i^- = \{v_1, v_2\}$, downstream roads in $\mathcal{N}_i^+ = \{j_1, j_2\}$, and roads in $\mathcal{I}_i = \{i, q\}$. In the distributed algorithm from Sect. IV-C, the i-th local subproblem (23) has variables indexed by roads in $\mathcal{S}_i$ only: primal variables $\bar{u}^{(i)}_{[p\in\mathcal{S}_i]}$ and dual variables $\lambda_i^{(i,p)}$, $\lambda_p^{(i,p)}$ for all $p \in \mathcal{S}_i$. For example, in this intersection, $\bar{u}^{(i)}_{[p\in\mathcal{S}_i]} = [\bar{u}^{(i)}_{v_1}, \bar{u}^{(i)}_{v_2}, \bar{u}^{(i)}_i, \bar{u}^{(i)}_q, \bar{u}^{(i)}_{j_1}, \bar{u}^{(i)}_{j_2}]$. Along iterations of the distributed algorithm, the i-th node exchanges messages with neighbors in $\mathcal{S}_i$, sharing its primal variables, so as to converge towards a solution satisfying the constraints (18c)-(18d): all local variables (from different local problems) concerning a same road converge to a same value; in this example, for road i, $\bar{u}_i^{(i)} = \bar{u}_i^{(v_1)} = \bar{u}_i^{(v_2)} = \bar{u}_i^{(q)} = \bar{u}_i^{(j_1)} = \bar{u}_i^{(j_2)}$.

B. Problem formulation in the distributed set-up

A common set-up in distributed optimization ([23]) deals with problems with a separable objective function and local constraints. Regarding the optimization presented in the previous section, we will show that we can rewrite the (centralized) problem (P3) as

$$\begin{aligned} \min_{\bar{u}} \quad & \sum_{i\in\mathcal{R}} g_i\big(\bar{u}_{[p\in\mathcal{S}_i]}\big) \\ \text{s.t. } & \bar{u}_{[p\in\mathcal{S}_i]} \in \mathcal{X}_i, \, \forall\, i \in \mathcal{R}, \end{aligned} \tag{17}$$

where each set $\mathcal{X}_i$ is compact and convex, and each function g_i is strictly convex. We recall that the notation $\bar{u}_{[p\in\mathcal{S}_i]}$ indicates a vector containing duty cycles of road i and of roads $p \in \mathcal{S}_i$.

The reason why this re-formulation is convenient lies in the existence of efficient algorithms to solve (17) in a distributed manner. More specifically, the separation into subproblems is naturally suggested by the structure of cost function and constraints. However, one more step is needed to formulate subproblems that do not have any variable in common. To this aim, we replace each variable $\bar{u}_r$ with several variables $\bar{u}_r^{(s)}$, one for each $s \in \mathcal{S}_r$, representing the local version known by node s of the variable $\bar{u}_r$. With this, the constraint $\bar{u}_{[p\in\mathcal{S}_i]} \in \mathcal{X}_i$ can be replaced by the constraint $\bar{u}^{(i)}_{[p\in\mathcal{S}_i]} \in \mathcal{X}_i$, which only involves local variables at node i. Then, additional constraints are needed, to ensure that all local versions of a same $\bar{u}_r$ are equal. This gives rise to the following problem:

$$\min_{[\bar{u}^{(i)}]_{i\in\mathcal{R}}} \quad \sum_{i\in\mathcal{R}} g_i\big(\bar{u}^{(i)}_{[p\in\mathcal{S}_i]}\big) \tag{18a}$$

$$\text{s.t. } \bar{u}^{(i)}_{[p\in\mathcal{S}_i]} \in \mathcal{X}_i, \, \forall\, i \in \mathcal{R}, \tag{18b}$$

$$\bar{u}_i^{(i)} = \bar{u}_i^{(p)}, \quad \forall\, i \in \mathcal{R}, \, \forall\, p \in \mathcal{S}_i \setminus i, \tag{18c}$$

$$\bar{u}_p^{(i)} = \bar{u}_p^{(p)}, \quad \forall\, i \in \mathcal{R}, \, \forall p \in \mathcal{S}_i \setminus i. \tag{18d}$$

The equivalence between (17) and (18) is guaranteed by the connectedness of the communication graph $\mathcal{G}$.

We will now describe the specific form of g_i and $\mathcal{X}_i$ that makes the one-step-ahead traffic optimization separable. Let g_i be defined as

$$\begin{aligned} & g_i\big(\bar{u}_{[p\in\mathcal{S}_i]}\big) = \\ & - k_{\text{ttd}} \min\Bigg\{ v_i\Big(\bar{\rho}_i(t) + \frac{T_s}{L_i}\Big(\sum_{j\in\mathcal{N}_i^-} \bar{u}_j \beta_{ji} \varphi_j^{\text{out}} - \bar{u}_i \varphi_i^{\text{out}}\Big)\Big), \\ & \qquad w_i\Big(\rho_i^{\max} - \bar{\rho}_i(t) - \frac{T_s}{L_i}\Big(\sum_{j\in\mathcal{N}_i^-} \bar{u}_j \varphi_j^{\text{out}} + \bar{u}_i \varphi_i^{\text{out}}\Big)\Big)\Bigg\} \\ & \qquad + k_{\text{bal}} \sum_{j\in\mathcal{N}_i^+} \Big(h_i(t)\bar{u}_{[i,v\in\mathcal{N}_i^-]} + \bar{\rho}_i(t) \\ & \qquad\qquad - h_j(t)\bar{u}_{[j,v\in\mathcal{N}_j^-]} - \bar{\rho}_j(t)\Big)^2 \\ & \qquad + \sum_{p\in\mathcal{S}_i} \frac{1}{|\mathcal{S}_p|}(\bar{u}_p - \bar{u}_p(t-T))^2, \end{aligned} \tag{19}$$

and $\mathcal{X}_i$ be defined as the set of all $\bar{u}_{[p\in\mathcal{S}_i]}$ satisfying

$$\begin{aligned} & \bar{u}_p \geq l_p \quad \forall\, p \in \mathcal{S}_i \\ & \bar{u}_p \leq 1 \quad \forall\, p \in \mathcal{S}_i \\ & \sum_{q\in\mathcal{I}_i} \bar{u}_q \leq 1. \end{aligned} \tag{20}$$

Based on the previous discussion, two remarks matter regarding definitions (19) and (20): first, every set $\mathcal{X}_i$ is compact and convex, since it is given by linear inequalities; second, every function g_i is stricly convex in every variable u_p, $p \in \mathcal{S}_i$. To see this, notice that each g_i is given by the sum of three terms: the first is convex because it is the composition of convex functions (notice that the $\min$ operator is multiplied by a negative constant); the second is convex because it is a quadratic form; the third is stricly convex because it is a quadratic form with a positive definite matrix (in particular, it is the identity matrix). Hence their sum is a stricly convex function.

Thanks to this definition of g_i and $\mathcal{X}_i$ and to their properties we have remarked, our problem is in the set-up specified by [23] whose steps we will now adapt to the traffic optimization problem.

It is important to notice that the function g_i can be defined by using only information from neighbors in $\mathcal{S}_i$. Indeed, as pointed out in Sect. III-D, it can be constructed as follows. First, each road i sends its density $\bar{\rho}_i(t)$ to its upstream neighbors in $\mathcal{N}_i^-$. Then, using its own density and the ones received from the downstream neighbors in $\mathcal{N}_i^+$, it computes φ_i^{out}. Road i sends φ_i^{out} to its downstream neighbors. Based on its own out-flow, and on the ones received from the upstream neighbors in $\mathcal{N}_i^-$, it can compute the TTD term of the cost function, as well as $h_i(t)$. In a final communication round, $h_i(t)$ is sent to upstream neighbors, and the balancing term is computed from the received messages from downstream. Finally, the regularization term involves $\bar{u}_p(t-T)$ for all $p \in \mathcal{S}_i$.

C. A distributed algorithm for the one-step-ahead traffic optimization

Following [23], we will now present a distributed algorithm to solve problem (18) that uses the well known technique of

dual decomposition. In (18), the objective function and the constraints (18b) are already separable. Therefore, the dual decomposition will consider the coupling constraints (18c) and (18d) only. To this aim, lagrangian multipliers are introduced, denoted $\lambda_i^{(i,p)}$ and $\lambda_p^{(i,p)}$ (for all $i \in \mathcal{R}$ and $p \in \mathcal{S}_i \setminus i$), associated with constraints (18c) and (18d), respectively. With some simple manipulations described in details in [23, Lemma 3.1], the partial lagrangian can be written as follows

$$L = \sum_{i\in\mathcal{R}} L_i = \sum_{i\in\mathcal{R}} \Big[g_i\big(\bar{u}^{(i)}_{[p\in\mathcal{S}_i]}\big) + \bar{u}_i^{(i)} \sum_{p\in\mathcal{S}_i\setminus i} \big(\lambda_i^{(i,p)} - \lambda_i^{(p,i)}\big) + \sum_{p\in\mathcal{S}_i\setminus i} \bar{u}_p^{(i)}\big(\lambda_p^{(i,p)} - \lambda_p^{(p,i)}\big)\Big]. \tag{21}$$

The dual decomposition algorithm is based on the iteration of two steps: the first is the minimization of the partial lagrangian, and the second is the update of the lagrangian multipliers λ via a gradient ascent update.

We now present the iterative algorithm, by illustrating the scheme that has to be executed by every subproblem i:

Initialization. Create local variables $\bar{u}_p^{(i)}$ for all $p \in \mathcal{S}_i$, and $\lambda_i^{(i,p)}$, $\lambda_p^{(i,p)}$ for all $p \in \mathcal{S}_i \setminus i$, and initialize them such that

$$\lambda_p^{(i,p)}(0) = -\lambda_p^{(p,i)}(0), \text{ for all } i \text{ and all } p \in \mathcal{S}_i \setminus i. \tag{22}$$

Set $k = 0$.

Primal update. Update primal variables $\bar{u}_p^{(i)}(k+1)$ by solving the local optimization problem:

$$\min_{\bar{u}^{(i)}_{[p\in\mathcal{S}_i]}} \quad g_i\big(\bar{u}^{(i)}_{[p\in\mathcal{S}_i]}\big) + 2\sum_{p\in\mathcal{S}_i\setminus i} \Big(\bar{u}_i^{(i)}\lambda_i^{(i,p)} + \bar{u}_p^{(i)}\lambda_p^{(i,p)}\Big) \quad \text{s.t. } \bar{u}^{(i)}_{[p\in\mathcal{S}_i]} \in \mathcal{X}_i. \tag{23}$$

Transmission. Send primal local variables to requiring neighbors $p \in \mathcal{S}_i$, and collect the most recent values of $\bar{u}_i^{(p)}$, $\bar{u}_p^{(p)}$ from them.

Dual update. Update the lagrangian multipliers: for every $p \in \mathcal{S}_i \setminus i$, set

$$\lambda_i^{(i,p)}(k+1) = \lambda_i^{(i,p)}(k) + \alpha\big(\bar{u}_i^{(i)}(k+1) - \bar{u}_i^{(p)}(k+1)\big),$$
$$\lambda_p^{(i,p)}(k+1) = \lambda_p^{(i,p)}(k) + \alpha\big(\bar{u}_p^{(i)}(k+1) - \bar{u}_p^{(p)}(k+1)\big). \tag{24}$$

Stop condition. If the solution has numerically stabilized stop, otherwise increment k and go to Primal update.

Notice that the primal update (23) is equivalent to minimization of L_i, the i-th term of the partial lagrangian given in (21). Indeed, if the inizialization of the lagrange multipliers satisfies (22), then the same property remains true along the gradient ascent updates (24), namely $\lambda_p^{(i,p)}(k) = -\lambda_p^{(p,i)}(k)$. As discussed in [23], the advantage of the expression of L_i used in (23), compared with the one in (21), is that the former only uses lagrangian multipliers locally known by node i, while the latter would require exchanging lagrangian multipliers with neighbors.

The convergence property of this algorithm is stated in the next proposition.

Proposition 3. *Let $\bar{u}^*$ be the unique optimal solution of the centralized optimization problem* (P3)*. There exists $\alpha^* > 0$ such that, if $0 < \alpha < \alpha^*$, then*

$$\lim_{k\to\infty} \bar{u}_i^{(i)} = \bar{u}_i^*,$$

for all $i \in \mathcal{R}$.

Proof. We have shown that $\mathcal{X}_i$ are compact and convex sets, and that functions g_i are stricly convex, thus satisfying the assumptions of [23, Theorem 3.3]. □

D. Comments

We have three remarks about the algorithm above. First, the initialization of local variables is subject to condition (22). There are two simple ways to satisfy this requirement:

- A simple initialization to zero values, every time the optimization starts;
- A zero initialization for the first optimization and a *warm start* for the following ones. Since in our set-up the optimization needs to be solved at the beginning of every traffic light cycle, is it possible to use the optimal values computed for previous cycle as initial condition for the new problem. Such a warm start choice is guaranteed to respect (22), because the update rule (24) preserves the same property. This approach can give practical benefits in terms of convergence speed.

In second place, it is important to notice that problems (23) can be constructed from the measured densities by exchanging messages within the neighborhood $\mathcal{S}_i$ only, as we explained in Sections III-D and IV-B.

Last, we would like to underline that the local optimization program stated in (23) can be rewritten as a convex (quadratic) program, using the same strategy outlined in Section III, with the introduction of auxiliary variables y_i (see Proposition 2). Therefore every local primal update is a convex program with a much smaller size than the centralized algorithm. The size of such local problems will not increase with the dimension of the network, only the number of subproblems will. This makes the algorithm suitable for large networks, in particular if an appropriate hardware infrastructure is available, i.e., every local problem can be solved in a different processing unit, and those units can communicate as prescribed by the communication graph $\mathcal{G}$.

V. Macroscopic numerical simulations

Optimization algorithms from Sections III and IV have been implemented and tested under various scenarios.

We recall that the goal of our algorithm is to decide about the duty cycle (i.e., the green split) of every traffic light in every intersection. The controller does not decide on the sequence of execution of such duty cycles. In other words, the order of the traffic lights within every cycle is given, and the objective of the optimal control is to assign the green duration of every traffic light.

In all our tests, since the cost function of the one-step-ahead control includes terms with different magnitudes (i.e., flows in the TTD and densities in the balancing measure),

we set the constant values k_{bal} and k_{ttd} such that both terms are normalized. In other words, every term contributing to the TTD, that is y_i, is normalized by the constant $\varphi_i^{\max}$, and every term $(\bar{\rho}_i - \bar{\rho}_j)^2$ is normalized by $\rho_i^{\max}$. Notice, also, that this numerical choice is straightfoward to implement in the distributed algorithm.

In this section, we report macroscopic simulations, run in MatLab environment, using software [37], [38] to solve the optimization problems. In Section V-A we present results about the convergence speed of the distributed implementation, while in Section V-B we show the improvement that the control strategy brings with respect to a best-practice policy for traffic lights, defined therein. All simulations in this section consider a Manhattan-like grid network, with same road parameters, split ratios, and external demands, as described in Section II-C.

We would like to underline that the averaged model is used only in the optimization algorithm, not in the simulations. At the beginning of every cycle, the optimal duty cycles (computed using the averaged model) are applied into the signalized CTM model, and at the beginning of the next cycle the density measurements are taken from this simulation. The performance of the algorithm is therefore evaluated in the signalized CTM system.

A. Convergence of the distributed traffic control

In order to estimate the speed of convergence of the distributed one-step-ahead control, we run a test bench of 2700 simulations, with different network sizes and initial densities. The network size varies among nine values, from 4 roads up to 180 roads. For each of these dimensions, we run 100 simulations, with initial state of every road pseudo-randomly generated every time. Such a test is repeated three times: first, the initial state is free-flow for all roads ($\rho(0) < \rho_c$); second, it is congested ($\rho(0) > \rho_c$); last, every road is initialized with a value that can be either free or congested ($0 < \rho(0) < \rho^{\max}$). The rationale is to test cases where roads have similar states (either free or congested) and where they have not (intuition would say that in the latter case a consensus requires more iterations).

The results are reported in the histograms in Figure 8. We declare that the distributed algorithm has converged when the change in its solution from one iteration to the next one is less (in absolute value) than 1e-3. The most important remark is that the maximum number of iterations before convergence is less than 30, in all our simulations. We were running our simulations in a sequential way on a laptop, but this algorithm is meant to be implemented in a parallel way in a distributed infrastructure, where each subproblem corresponds to a separate hardware (e.g., each traffic light is integrated with a microprocessing unit). We can estimate the total computation time over such a parallel infrastructure, by considering that in our tests an iteration takes around 0.05 seconds for each subproblem. This means that the total computation time for the parallel infrastructure would be always less than 0.15 s, much lower than the sampling time $T_s = 15$ s.

Time needed for communication should be taken into account too; however, in our algorithm communication exists only between roads physically close, hence the time needed for all iterations results to be comparable to the time needed by the centralized algorithm, in which the solution, once it has been computed, needs to be sent from the single computing machine to roads that are also very far.

We also notice that the number of iterations before convergence varies with the variability of the free-congested mode over the network. If all cells have the same mode (all free or all congested), then a smaller number of iterations is needed, at most 18.

B. Comparison with a best-practice traffic lights control

Most cities apply a simple choice for green time of traffic lights, i.e., selecting fixed duration based on the known turning proportions and the statistical knowledge of external demands, as it is implemented in the earlier traffic lights control policies, e.g., [6]. In this paper, we compare our feedback strategy with a fixed best-pratice policy based on the same concepts: we collect data from the simulated network and we use it to compute the mean density experienced within every road during the simulation. Given this information, fixed duty cycles are assigned proportionally to the mean densities.

Representative results are reported in Figure 9, where we consider a network with 40 roads. We show, for every road and for every time step, the distance between the road density $\rho_i(t)$ and the critical density ρ_c; recall that ρ_c maximizes the local flow. Also, we show a graphical visualization of the two optimized indexes (TTD and density balancing), from which the improvement can be seen. Moreover, we show how another well-known traffic index, the service of demand, behaves during the system evolution. This index represents the ability to cope with external demands, allowing vehicles to enter the network; it is defined [31] as the sum of entering flows

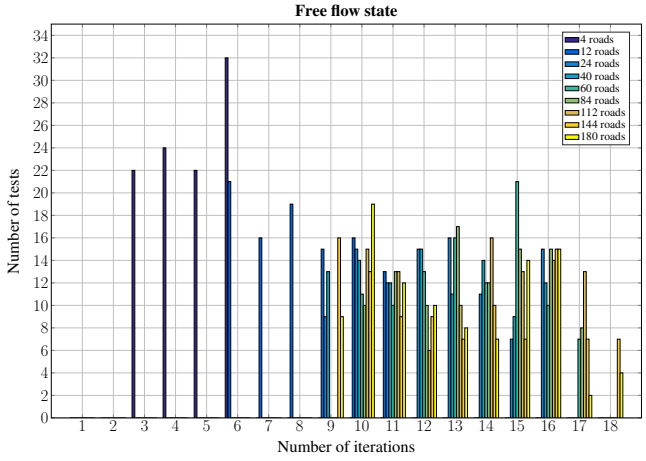
$$\text{SoD}(t) = \sum_{k=0}^{\lfloor t/T_s \rfloor} \sum_{i \in \mathcal{R}^{\text{in}}} \varphi_i^{\text{in}}(kT_s),$$

We remark that, even though this index was not included in the optimization process, an improvement is also achieved in its respect.

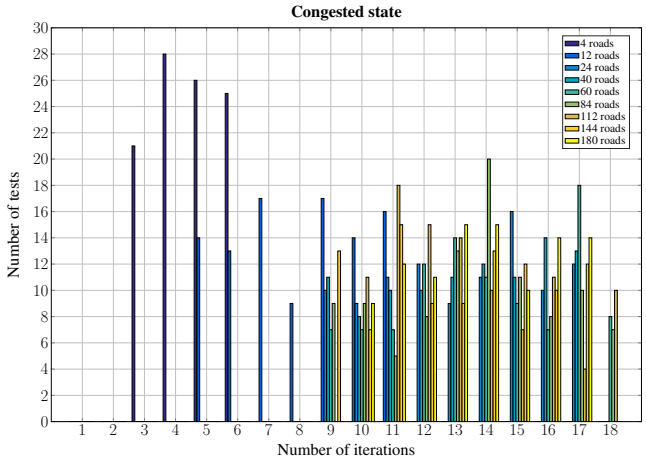
VI. Microscopic experiments

The algorithms illustrated in this paper have also been tested in a commercial traffic microsimulator (Aimsun, [24]). Such a software simulates traffic behavior within roads by implementing a microscopic model for each and every vehicle. This allows us to see how our algorithm behaves when applied to a model very different from the macroscopic model used for the optimization, thus showing its potential for application on real world.

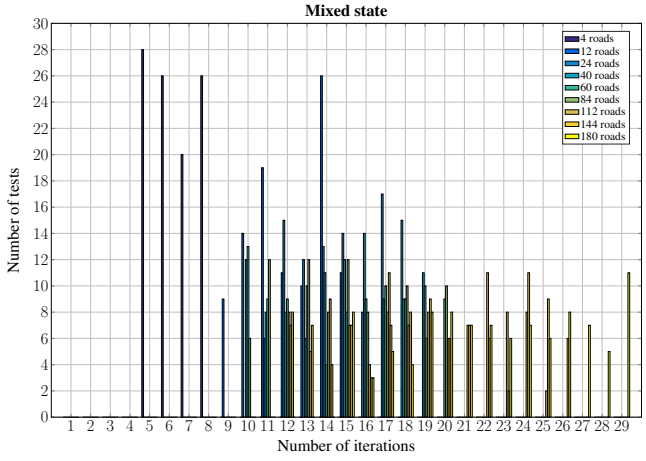
For our microscopic experiments, we run simulations in which, while Aimsun software keeps running, the one-step ahead optimal control is computed and is then injected into Aimsun, that will continue the simulation with the new values for the green times. More precisely, we use the method proposed in this paper, but relaxing the assumption that all traffic lights have the same cycle length. Indeed, in our realistic



(a)

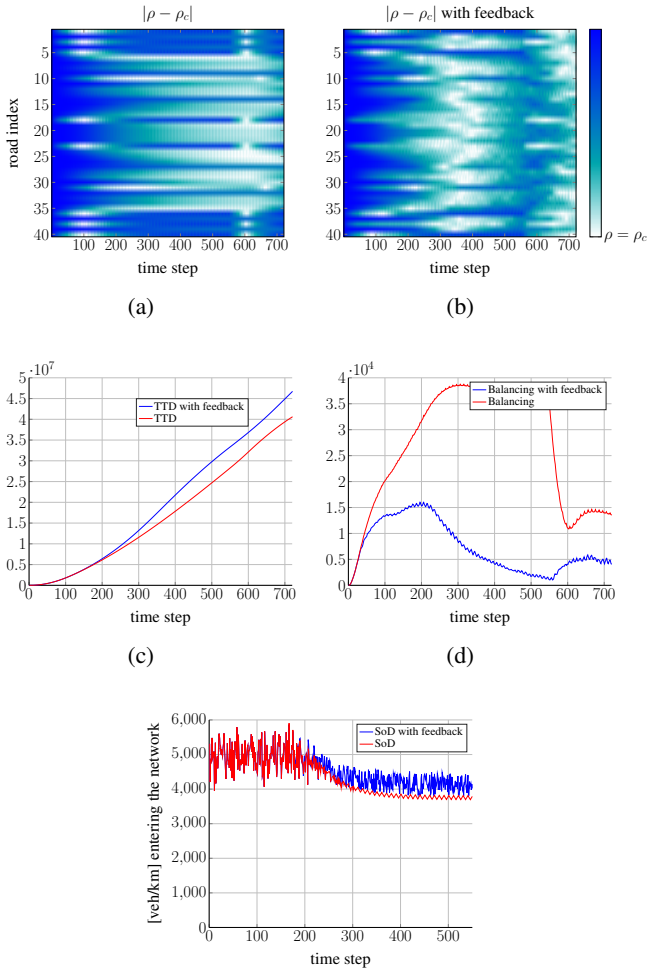


(b)



(c)

Fig. 8: Results of the test bench on convergence speed of the distributed algorithm.



(a) (b) (c) (d) (e)

Fig. 9: Comparison between the best-practice control and the feedback strategy proposed in this paper. Figures 9a and 9b show the distance $|\rho - \rho_c|$ for the two cases; the critical density ρ_c can be interpreted as the ideal working point, maximizing flow. The improvement achieved in terms of TTD is shown in Figure 9c, and the improvement in terms of balancing in Figure 9d. In Figure 9e we show the improvement over another traffic index, the service of demand (SoD), that the algorithm obtains even though such an index is not explicitly part of the cost function of the optimization problem.

scenario, cycle length depends on the size of the intersection, and we have four groups of traffic lights, corresponding to cycle lengths of 45, 60, 90, and 120 seconds; traffic lights in a same group have a common starting time. At a time t when one group begins a cycle, the optimization problem is solved across all the network, but the resulting optimal green times are applied only in this group, while all other traffic lights keep applying their previously-computed solution, until the end of their cycles.

For comparison, other simulations are run, where traffic lights are instead controlled by a best practice rule. This is a commonly adopted policy, where duty cycles are time-invariant, and are assigned according to the split ratios and to the historical demands [6]. To implement this policy in our simulations, we have collected 'historical' data, running the simulator without traffic lights, and then assigned duty cycles

proportionally to the mean densities.

To generate the scenarios, we used a Aimsun model that reflects the downtown portion of Grenoble, France, in every major aspect (see Figure 10). The computation of the optimal control was done in the Python language, as Aimsun supports Python scripting through its API. The solver [38] has been used, the same as in the MatLab simulations in Section V.

We used real flows measurements (obtained from the loop detectors installed in the city) in order to improve the Aimsun model, as it follows.

Some of these loop detectors are placed near the boundary of the town, and from them we measured the external input values, that are also injected into the Aimsun model (in terms of vehicles per hour). In Figure 11 we report some examples of these data profiles.

The other detectors, placed more towards the center of the city, have been used to reproduce realistic values of split ratios into Aimsun: by measuring the flow in a road from a detector, and the input flows in the downstream neighbors of the same road, we could compute the split ratios for almost all pairs of roads. In the few cases where this was not possible (due to the absence of a detector) we assigned the splits proportionally to the size of the downstream roads (i.e., to the number of lanes): for example, in an intersection with three exiting roads with 1, 2 and 3 lanes respectively, they would be assigned split ratios equal to 0.166, 0.332 and 0.502, respectively, from each of the roads entering the same intersection. Such a rule is meant to reproduce the fact that larger streets are used by more vehicles.

Finally, we extracted the CTM parameters from the Aimsun user interface, after we imported the (open street) map of the model.

We have selected a typical day and set up the simulation as to represent the stream of vehicles corresponding to the morning peak-hours, from 7 a.m. to 10 a.m. Each simulation is then run for three (simulated) hours.

We report about experiments conducted in two different scenarios. In both, the entering flows are the same, but in the second scenario we force the software to create incidents which give increasing congestions in some areas. At the end of every simulation, Aimsun computes microscopic traffic measures that we used to evaluate the performance of the proposed strategy. Such results are shown in Table II. We remark that every microscopic index is improved by using the presented control algorithm: travelled distance is increased, in average, of around 13%; travel time is reduced by more than 17% in the first scenario, and almost 20% in the second; the mean queue length (summed over all roads in the network) is reduced by 10%; last, the stop time (computed as the time the vehicles were stopped, per kilometer) is reduced by 21% in the first scenario, and by 19% in the second.

A demo of these experiments is available online [39].

TABLE II: Evaluation of the one-step-ahead optimal control (OSA-OC) against the best-practice policy (BP), via microscopic experiments and resulting traffic indexes computed by the microscopic simulator Aimsun.

Index	Scenario 1		Scenario 2	
	BP	OSA-OC	BP	OSA-OC
Travelled distance [km]	23396	26471	19772	17003
Travel time [h]	1775	1462	1955	1583
Mean queue [veh]	496	441	627	564
Stop time [sec/km]	123	97	172	139

Fig. 10: From top left, clockwise: a map of Grenoble downtown; the corresponding Aimsun model used in the microscopic experiments, with markers for the detector positions that correspond to the inflows of the model (see also Figure 11); the directed graph describing this road network. A demo of our experiments is available online [39].

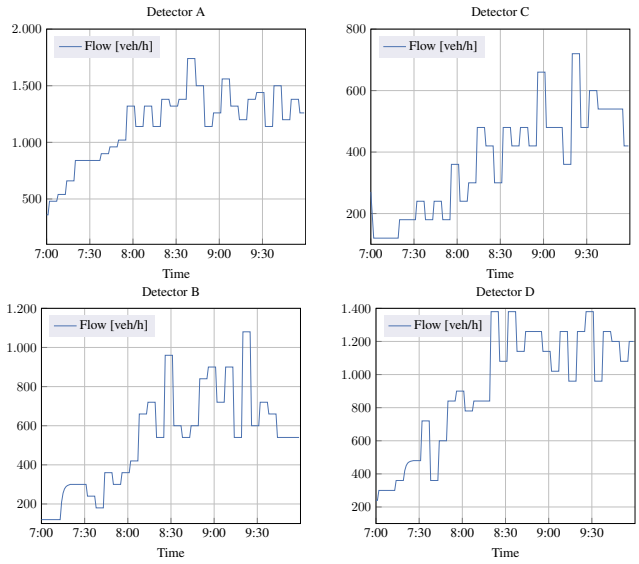


Fig. 11: Some of the real flow profiles that we used as input for the Aimsun model. The plots correspond to flows from loop detectors in positions A, B, C, D in Figure 10.

VII. Conclusions

In this work we have proposed a distributed algorithm for optimal assignment of traffic lights in urban networks. The proposed approach is based on the decomposition of a convex centralized optimization, based on a one-step-ahead prediction,

that achieves optimal performance with respect to the total travel distance and the density balancing, two well-known traffic indexes. We formally prove that the distributed scheme converges to the optimal solution of the centralized problem. The designed control scheme is adaptive to fast changes in the enviroment. Moreover, thanks to the distributed set-up, the algorithm satisfies the scalability requirement needed for large-scale networks, since the problem is split in several subproblems whose sizes do not increase with the network size.

We have validated the work also through numerical simulations, both in macroscopic and in microscopic environment. The results of these numerical tests show that the distributed procedure converges efficiently to the optimal solution. We also measured the improvement that it offers over a best-practice policy for traffic lights, not only with respect to the the considered objectives, but also to other traffic indexes such as travel time, mean queue length and stop time, that are consistently improved.

VIII. Acknowledgement

The authors would like to thank engineer Rohit Singal who created the Aimsun model.

This work has been partly funded by the EU FP7 project SPEEDD (619435) and by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement 694209).

References

[1] X. He and L. Zhang, "Economic impact of highway investment at the metropolitan level: Empirical analysis with consideration for induced demand and induced supply," *Transportation Research Board Annual Meeting*, 2013.

[2] M. Sweet, "Traffic Congestion's Economic Impacts: Evidence from U.S. Metropolitan Regions," *Urban Studies*, vol. 51, no. 10, pp. 2088–2110, 2014.

[3] G. De Nunzio, G. Gomes, C. Canudas-de-Wit, R. Horowitz, and P. Moulin, "Arterial bandwidth maximization via signal offsets and variable speed limits control," *54th IEEE Conference on Decision and Control*, pp. 5142–5148, 2015.

[4] E. Lovisari, G. Como, A. Rantzer, and K. Savla, "Stability analysis and control synthesis for dynamical transportation networks," *arXiv preprint arXiv:1410.5956*, 2014.

[5] Q. Ba, K. Savla, and G. Como, "Distributed optimal equilibrium selection for traffic flow over networks," *54th IEEE Conference on Decision and Control*, pp. 6942–6947, 2015.

[6] D. Robertson, "TRANSYT method for area traffic control," *Traffic Engineering Control*, vol. 11, no. 6, pp. 276–281, 1969.

[7] P. Hunt, D. Robertson, R. Bretherton, and R. Winton, "SCOOT - A traffic responsive method of coordinating signals," Transport Research Laboratory, Tech. Rep., 1981. [Online]. Available: https://trl.co.uk/reports/LR1014

[8] K. Aboudolas, M. Papageorgiou, and E. Kosmatopoulos, "Store-and-forward based methods for the signal control problem in large-scale congested urban road networks," *Transportation Research Part C: Emerging Technologies*, vol. 17, no. 2, pp. 163–174, 2009.

[9] K. Aboudolas, M. Papageorgiou, A. Kouvelas, and E. Kosmatopoulos, "A rolling-horizon quadratic programming approach to the signal control problem in large scale congested urban road networks," *Transportation Research Part C: Emerging Technologies*, vol. 18, no. 5, pp. 680 – 694, 2010.

[10] P. Varaiya, "Max pressure control of a network of signalized intersections," *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 177–195, 2013.

[11] T. Le, P. Kovács, N. Walton, H. L. Vu, L. L. H. Andrew, and S. Hoogendoorn, "Decentralized signal control for urban road networks," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 431–450, Sept. 2015.

[12] C. F. Daganzo, "The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory," *Transportation Research Part B: Methodological*, vol. 28, no. 4, pp. 269–287, 1994.

[13] A. Muralidharan and R. Horowitz, "Optimal control of freeway networks based on the link node cell transmission model," *IEEE American Control Conference*, pp. 5769–5774, 2012.

[14] G. Como, E. Lovisari, and K. Savla, "Convexity and robustness of dynamic traffic assignment and freeway network control," *Transportation Research Part B: Methodological*, vol. 91, pp. 446–465, 2016.

[15] G. Gomes and R. Horowitz, "Optimal freeway ramp metering using the asymmetric cell transmission model," *Transportation Research Part C: Emerging Technologies*, vol. 14, no. 4, pp. 244–262, 2006.

[16] D. Pisarski, "Collaborative ramp metering control: Application to Grenoble south ring," Ph.D. dissertation, Univ. Grenoble Alpes, 2015.

[17] D. Pisarski and C. Canudas-de-Wit, "Optimal balancing of road traffic density distributions for the cell transmission model," *IEEE 51st Annual Conference on Control and Decision*, pp. 6969–6974, 2012.

[18] K. Han, V. V. Gayah, B. Piccoli, T. L. Friesz, and T. Yao, "On the continuum approximation of the on-and-off signal control on dynamic traffic networks," *Transportation Research Part B: Methodological*, vol. 61, pp. 73–97, 2014.

[19] C. Canudas-de-Wit, "An Average Study of the Signalized Cell Transmission Model," preprint. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01635539/document

[20] S. Ukksuri, G. Ramadurai, and G. Patil, "A robust transportation signal control problem accounting for traffic dynamics," *Computers and Operations Research*, vol. 37, no. 5, pp. 869–879, 2010.

[21] H. Khalil, *Nonlinear systems*. Prentice Hall, 2002.

[22] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[23] R. Carli and G. Notarstefano, "Distributed partition-based optimization via dual decomposition," in *52nd IEEE Conference on Decision and Control*, 2013, pp. 2979–2984.

[24] TSS - Transportation Simulation Systems, *Aimsun 8 Dynamic Simulators Users' Manual*, 2015.

[25] M. J. Lighthill and G. B. Whitham, "On kinematic waves II. A theory of traffic flow on long crowded roads," *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 229, no. 1178, pp. 317–345, 1955.

[26] P. I. Richards, "Shock waves on the highway," *Operations Research*, vol. 4, no. 1, pp. 42–51, 1956.

[27] C. F. Daganzo, "The cell transmission model, part II: Network traffic," *Transportation Research Part B: Methodological*, vol. 29, no. 2, pp. 79–93, 1995.

[28] J. Lebacque, "The Godunov scheme and what it means for first-order traffic models," *13th International Symposium on Transportation and Traffic Theory*, pp. 647–77, 1996.

[29] J. Little, "The synchronization of traffic signals by mixed integer linear programming," *Operation Research*, vol. 14, no. 4, pp. 568–594, 1966.

[30] P. Grandinetti, F. Garin, and C. Canudas-de-Wit, "Towards scalable optimal traffic control," *54th IEEE Conference on Decision and Control*, pp. 2175–2180, 2015.

[31] P. Grandinetti, C. Canudas-de-Wit, and F. Garin, "An efficient one step ahead optimal control for urban signalized traffic networks based on an averaged Cell Transmission Model," *14th European Control Conference*, pp. 3478–3483, 2015.

[32] T. Le, H. L. Vu, Y. Nazarathy, Q. B. Vo, and S. Hoogendoorn, "Linear-quadratic model predictive control for urban traffic networks," *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 498–512, Nov. 2013.

[33] S. Samaranayake, J. Reilly, W. Krichene, J. B. Lespiau, M. L. Delle Monache, P. Goatin, and A. Bayen, "Discrete-time system optimal dynamic traffic assignment (SO-DTA) with partial control for horizontal queuing networks," *IEEE American Control Conference*, pp. 663–670, 2015.

[34] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[35] D. Richert and J. Cortés, "Robust distributed linear programming," *IEEE Trans. Automatic Control*, vol. 60, no. 10, pp. 2567–2582, 2015.

[36] M. Burger, G. Notarstefano, F. Bullo, and F. Allgower, "A distributed simplex algorithm for degenerate linear programs and multi-agent assignments," *Automatica*, vol. 8, no. 9, pp. 2298–2304, 2012.
[37] J. Löfberg, "YALMIP : A Toolbox for Modeling and Optimization in MATLAB," *Proceedings of the CACSD Conference*, 2004.
[38] MOSEK ApS, *The MOSEK optimization toolbox for MATLAB manual. Version 7.1.*, 2015. [Online]. Available: http://docs.mosek.com/7.1/toolbox/index.html
[39] "Grenoble Traffic Lab – Aimsun demo," http://gtl.inrialpes.fr/aimsun-control-demo-accident/index.html.

Pietro Grandinetti was born in Cosenza (Italy) in 1990. He received his B.S. degree in Computer Engineering and MS degree in Automation Engineering from Università degli studi della Calabria (Italy), in 2011 and 2013 respectively, and his PhD in Automation Engineering from Université Grenoble Alpes (France) in 2017. During his PhD studies, he has been interested in traffic networks modelling and distributed control.

Federica Garin (M'16) is a researcher ('chargée de recherche') with the NeCS team at INRIA and GIPSA-lab, Grenoble (France). She received her B.S., M.S., and Ph.D. degrees in Applied Mathematics from Politecnico di Torino (Italy) in 2002, 2004, and 2008, respectively. She was a post-doctoral researcher at Università di Padova (Italy) in 2008 and 2009, and at INRIA in 2010. She is an Associate Editor in the IEEE-CSS Conference Editorial Board and in the European Control Association (EUCA) Conference Editorial Board. Her current research interests are in distributed algorithms and network systems.

Carlos Canudas-de-Wit (F'16) was born in Villahermosa, Mexico, in 1958. He received his B.S. degree in electronics and communications from the Monterrey Institute of Technology and Higher Education, Monterrey, Mexico, in 1980, and the M.S. and Ph.D. degrees in automatic control from the Department of Automatic Control, Grenoble Institute of Technology, Grenoble, France, in 1984 and 1987, respectively. He is currently 'Directeur de recherche' (Senior researcher) with CNRS in Grenoble, France, where he is the leader of the NeCS team, a joint team of GIPSA-lab (CNRS) and INRIA, on Networked Controlled Systems. Dr. Canudas-de-Wit is an IFAC Fellow and an IEEE Fellow. He has been associate editor of the IEEE Transactions on Automatic Control, of Automatica, and of the IEEE Transactions on Control Systems Technology. He is currently an Associate Editor of the Asian Journal of Control and of the IEEE Transactions on Control of Network Systems. He served as the President of the European Control Association from 2013 to 2015, and for the IEEE Board of Governors of the Control System Society from 2011 to 2014. He holds the ERC Advanced Grant Scale-FreeBack from 2016 to 2021.